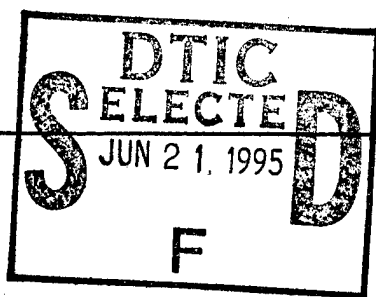| REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 12-19-94	3. REPORT TYPE AND DATES COVERED Final Project Report 1992-1994	

4. TITLE AND SUBTITLE Design and Analysis of Lossless and Lossy Data Compression Methods with Applications to Communication and Caching	5. FUNDING NUMBERS $112,774 0304/FS

6. AUTHOR(S)

Jeffrey S. Vitter

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Duke University Department of Computer Science Box 90129 Durham, NC 27708-0129	8. PERFORMING ORGANIZATION REPORT NUMBER AFOSR-TR. 9 5 - 0 3 0 3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research /NM 110 Duncan Avenue, Suite B115 Bolling AFB Washington, DC 20332-0001	10. SPONSORING/MONITORING AGENCY REPORT NUMBER F49620-92-J-0515

11. SUPPLEMENTARY NOTES

The views, opinions and/or finds contained in this report are those of the author(s) and should not be construed as an official Air Force Office of Scientific Research position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited.**	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Please see attached.

DTIC ELECTED
JUN 2 1, 1995
F

DTIC QUALITY INSPECTED 5

14. SUBJECT TERMS			15. NUMBER OF PAGES 7
			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Project Report

AFOSR Grant No. F49620-92-J-0515
Brown University
(subcontracted to Duke University)

1992–1994

Jeffrey S. Vitter
Lehrman Professor and Chair
Department of Computer Science
Duke University
Durham, NC 27708-0129

Summary of Progress

Our work in this one-year project has been very successful. It has revolved around the following three important areas dealing data compression and its applications:

- the design and analysis of sophisticated methods for prediction based on data compression techniques, with applications to prefetching, caching, and locality management.

- fast, practical, and code-efficient implementations of arithmetic coding and other coding methods, for use in text and image compression.

- new methods for choosing motion vectors yielding substantially better rate-distortion tradeoffs for video compression in videoconferencing applications.

Technology Transfer

Duke University recently filed a patent application [15] for the work on prediction.

Optimal Prediction via Data Compression

Caching and prefetching are important mechanisms for speeding up access time to data on secondary storage. In their *FOCS '91* paper, Prof. Vitter and graduate student P. Krishnan develop an optimal universal prefetcher in terms of fault ratio, with particular applications to large-scale databases and hypertext systems. The algorithms are novel in that they are based on data compression techniques that are both theoretically optimal and good in practice. They show for powerful models such as Markov sources and mth-order Markov sources that the page fault rates are optimal in the limit for almost all sequences of page accesses.

An important issue that affects response time performance in current OODB and hypertext systems is the I/O involved in moving objects from slow memory to cache. A promising way to tackle this problem is to use *prefetching*, in which we predict the user's next page requests and get those pages into cache in the background. Current databases perform limited prefetching using techniques derived from older virtual memory systems. A novel idea of using data compression techniques for prefetching was recently advocated by Vitter in Krishnan in which the prefetchers

based on the Lempel-Ziv data compressor (the UNIX *compress* command) were shown theoretically to be optimal in the limit.

In our work [3], we analyze the *practical* aspects of using data compression techniques for prefetching. We adapt three well-known data compressors to get three simple, deterministic, and universal prefetchers. We simulate our prefetchers on sequences of page accesses derived from the OO1 and OO7 benchmarks and from CAD applications, and demonstrate significant reductions in fault-rate. We examine the important issues of cache replacement, size of the data structure used by the prefetcher, and problems arising from bursts of "fast" page requests (that leave virtually no time between adjacent requests for prefetching and book keeping). We conclude that prediction for prefetching based on data compression techniques holds great promise.

In terms of mathematical analysis of prefetching, a different approach than that used for caching must be followed. Unlike other online problems, prefetching cannot admit a competitive analysis, since the optimal offline prefetcher incurs no cost when it knows the future page requests. Previous analytical work on prefetching by Vitter and Krishnan consisted of modeling the user as a probabilistic Markov source.

In [14] we look at the much stronger form of worst-case analysis and derive a randomized algorithm that we prove analytically *converges almost surely to the optimal fault rate in the worst case for every sequence of page requests* with respect to the important class of finite state prefetchers. In particular, we make no assumption about how the sequence of page requests is generated. This analysis model can be looked upon as a generalization of the competitive framework, in that it compares an online algorithm in a worst-case manner over all sequences against a powerful yet non-clairvoyant opponent. We simultaneously achieve the computational goal of implementing our prefetcher in optimal constant expected time per prefetched page, using the optimal dynamic discrete random variate generator of Matias, Vitter, and Ni in *Proceedings of the 4th Annual SIAM/ACM Symposium on Discrete Algorithms, Austin, TX, January 1993, 361–370.*

We are currently looking at applications of these prediction techniques to other locality management applications, such as power management for mobile computers.

Arithmetic Coding, Text Compression, and Image Compression

Arithmetic coding, in conjunction with a suitable probabilistic model, can provide nearly optimal data compression. In [5] analysis of arithmetic coding .], Prof. Vitter and Paul Howard (formerly graduate student and postdoctoral assistant) analyze the effect that the model and the particular implementation of arithmetic coding have on the code length obtained. They show that adaptive models give the same code length as semi-adaptive decrementing models. Periodic scaling is often used in arithmetic coding implementations to reduce time and storage requirements; it also introduces a recency effect that can further affect compression. They introduce the notion of "weighted entropy" and use it to characterize in an elegant way the effect that periodic scaling has on the code length. They also give a rigorous proof that the coding effects of rounding scaled weights, using integer arithmetic, and encoding end-of-file are negligible.

Prof. Vitter and Paul Howard in [13, 12] provide the basis of a fast, space-efficient, approximate arithmetic coder with only minimal loss of compression efficiency. The coder, called *quasi-arithmetic coding*, is based on the replacement of arithmetic coding by fast, carefully arranged table lookups, coupled with a new deterministic probability estimation scheme.

They give a detailed algorithm for fast text compression in [13] based partly on quasi-arithmetic coding. The algorithm, related to the PPM methods, which are the state-of-the-art methods for maximum text compression, simplifies the modeling phase by eliminating the *escape* mechanism, and speeds up coding by using a combination of *quasi-arithmetic coding* and Rice coding. They provide details of the use of quasi-arithmetic code tables, and analyze their compression perfor-

2

mance. Our *Fast PPM* method is shown experimentally to be almost twice as fast as the PPMC method, while giving comparable compression.

In [9] Prof. Vitter and Paul Howard give a new paradigm for lossless image compression, with four modular components: pixel sequence, prediction, error modeling, and coding. They present two new methods (called *MLP* and *PPPM*) for lossless compression, both involving linear prediction, modeling prediction errors by estimating the variance of a Laplace distribution, and coding using arithmetic coding applied to precomputed distributions. MLP is both progressive and parallelizable. The new methods compress high-resolution images significantly better than do other lossless methods, including the lossless mode of JPEG.

MLP compression is improved considerably in [8] by a preliminary method for error modeling using the *variability index*, which provides accurate models for pixel prediction errors without requiring explicit transmission of the models. The variability index can also be used to show that prediction errors do not always follow the Laplace distribution, as is commonly assumed; replacing the Laplace distribution with a more general distribution can improve compression. A fast PRAM implementation of MLP is proposed in [7] by the same authors.

Prof. Vitter and Paul Howard used the results of their modeling studies with the MLP method to develop an extremely fast lossless image compressor called FELICS, which compresses as efficiently as the JPEG lossless mode and runs five times faster [10] Work is proceeding on a progressive variant based on a hierarchical pixel sequence [11] The same authors applied their modeling expertise to address the problem of speed of the PPM family of text compressors, which are currently the most effective methods in terms of the amount of compression achieved. The Ziv-Lempel methods (such as the UNIX *compress* command) are much faster but do not compress as well. The new PPMD method offers state-of-the-art compression at double the speed of previous PPM methods [12].

Video Compression via Motion Compensation

In [4] we compare methods for choosing motion vectors for motion-compensated video compression. Our primary focus is on videophone and videoconferencing applications, where very low bit rates are necessary, where the motion is usually limited, and where the frames must be coded in the order they are generated. We provide evidence, using established benchmark videos of this type, that choosing motion vectors to minimize codelength subject to (implicit) constraints on quality yields substantially better rate-distortion tradeoffs than minimizing notions of prediction error. We illustrate this point using an algorithm within the $p \times 64$ standard. We show that using quadtrees to code the motion vectors in conjunction with explicit codelength minimization yields further improvement. We describe a dynamic-programming algorithm for choosing a quadtree to minimize the codelength. Current research is aimed at heuristics for speeding up the processing time and use of similar ideas for gaining improvements in static image compression.

Machine Learning

In [1] we introduce a new technique which enables a learner without access to hidden information to learn nearly as well as a learner with access to hidden information. We apply our technique to solve an open problem of Maass and Turán. We describe analogous results for two generalizations of this model to function learning, and apply those results to bound the difficulty of learning in the harder of these models in terms of the difficulty of learning in the easier model. We bound the difficulty of learning unions of k concepts from a class F in terms of the difficulty of learning F. We bound the difficulty of learning in a noisy environment for deterministic algorithms in terms of the difficulty of learning in a noise-free environment. We apply a variant of our technique to develop an algorithm transformation that allows probabilistic learning algorithms to nearly optimally cope with noise. A second variant enables us to improve a general lower bound of Turán for the PAC-

learning model (with queries). Finally, we show that logarithmically many membership queries never help to obtain computationally efficient learning algorithms.

In [2], we consider the problem of learning real-valued functions from random examples when the function values are corrupted with noise. With mild conditions on independent observation noise, we provide characterizations of the learnability of a real-valued function class in terms of a generalization of the Vapnik-Chervonenkis dimension, the fat-shattering function, introduced by Kearns and Schapire. We show that, given some restrictions on the noise, a function class is learnable in our model if and only if its fat-shattering function is finite. With different (also quite mild) restrictions, satisfied for example by gaussian noise, we show that a function class is learnable from polynomially many examples if and only if its fat-shattering function grows polynomially. We prove analogous results in an agnostic setting, where there is no assumption of an underlying function class.

References in Research Summary

1. P. Auer and P. M. Long. "Simulating access to hidden information while learning," *Proceedings of the 26th Annual ACM Symposium on the Theory of Computation*, 1994.

2. P. L. Bartlett, P. M. Long and R. C. Williamson. "Fat Shattering and the Learnability of Real-valued Functions," *Proceedings of the 1994 Workshop on Computational Learning Theory*.

3. K. M. Curewitz, P. Krishnan, and J. S. Vitter. "Practical Prefetching via Data Compression," *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (SIGMOD '93)*, Washington, D. C, May 1993.

4. D. T. Hoang, P. M. Long, and J. S. Vitter, "Explicit Bit Minimization for Motion-Compensated Video Coding," *Proceedings of the 1994 IEEE Data Compression Conference (DCC '94)*, Snowbird, UT, March 1994.

5. P. G. Howard and J. S. Vitter. "Analysis of Arithmetic Coding for Data Compression," invited paper in special issue on data compression for images and texts in *Information Processing and Management*, 1992, 749–763.

6. P. G. Howard and J. S. Vitter. "New Methods for Lossless Image Compression Using Arithmetic Coding," invited paper in special issue on data compression for images and texts in *Journal of Information Processing and Management*, 28(6), 1992, 765–779.

7. "Parallel Lossless Image Compression Using Huffman and Arithmetic Coding," in preparation. A shortened version appears in *Proceedings of the 1992 IEEE Data Compression Conference (DCC '92)*, Snowbird, UT, March 1992, 299–308.

8. P. G. Howard and J. S. Vitter. "Error Modeling for Hierarchical Lossless Image Compression," in preparation. A shortened version appears in *Proceedings of the 1992 IEEE Data Compression Conference (DCC '92)*, Snowbird, UT, March 1992, 269–278.

9. P. G. Howard and J. S. Vitter. "New Methods for Lossless Image Compression Using Arithmetic Coding," invited paper in special issue on data compression for images and texts in *Information Processing and Management*, 1992, 765–779.

10. P. G. Howard and J. S. Vitter. "Fast and Efficient Lossless Image Compression," *Proceedings of the 1993 IEEE Data Compression Conference (DCC '93)*, Snowbird, UT, April 1993.

11. P. G. Howard and J. S. Vitter. "Fast Progressive Lossless Image Compression," *Proceedings of the 1994 IST/SPIE International Symposium on Electronic Imaging Science and Technology*, San Jose, CA, February 1994.

12. P. G. Howard and J. S. Vitter. "Arithmetic Coding for Data Compression," *Proceedings of the IEEE*, 82(6), June 1994.

13. P. G. Howard and J. S. Vitter. "Design and Analysis of Fast Text Compression Based on Quasi-Arithmetic Coding," to appear in *Journal of Information Processing and Management*.

14. P. Krishnan and J. S. Vitter. "Optimal Prediction for Prefetching in the Worst Case," *Proceedings of the 5th Annual SIAM/ACM Symposium on Discrete Algorithms (SODA '94)*, Alexandria, VA, January 1994.

15. J. S. Vitter, K. M. Curewitz, and P. Krishnan, "Online Background Predictors and Prefetchers," Serial No. 08/246,600, filed May 20, 1994, patent pending.

Number of researchers supported

Faculty (including the P.I.): 1
Postdocs: 2 (Prof. Philip M. Long and Dr. Paul G. Howard)
Graduate Students: 1 (Mr. Dzung T. Hoang)

Professional honors and participation

Prof. Jeffrey S. Vitter:

- elected Fellow of Institute of Electrical and Electronics Engineers (IEEE).

- appointed to the Gilbert, Louis, and Edward Lehrman Professorship of Computer Science at Duke University.

- Program Committees:
  - 1993 IEEE Data Compression Conference (DCC '93), Snowbird, UT, March 1993.
  - 1993 Workshop on Algorithms and Data Structures (WADS '93), Montreal, Canada, August 1993.
  - 1994 IEEE Data Compression Conference (DCC '94), Snowbird, UT, March 1994.
  - 7th Annual ACM Conference on Computational Learning Theory (COLT '94), New Brunswick, NJ, July 1994.
  - Conference Vice-Chair, Member of Program Committee, and Session Chair, 6th IEEE International Conference on Tools with Artificial Intelligence (TAI '94), New Orleans, LA, November 1994.

- Invited talks:
  - "A Theory for Memory-Based Learning," Massachusetts Institute of Technology, Cambridge, MA.
  - "Paradigms for Optimal Sorting with Multiple Disks and Memory Hierarchies," Plenary address at 7th Maryland Theoretical Computer Science Day, Johns Hopkins University, Baltimore, MD.
  - "The Design of Lossless Image Compression Systems," Workshop on Data and Image Compression Needs and Uses in the Scientific Community, Center of Excellence in Space Data and Information Sciences, NASA Goddard Space Flight Center, Greenbelt, MD.

- "Locality, Dynamic, and Prediction Issues in DIS," Panel Member, ARO Workshop on Virtual, Distributed Interactive Simulation, Research Triangle Park, NC.
- "Obstacles in the Implementation of Parallel Algorithms," Panel Member, Workshop on Parallel I/O and Databases, Dartmouth Institute for Advanced Graduate Studies (DAGS 93), Hanover, NH.
- "Load Balancing Paradigms for Optimal Use of Parallel Disks and Parallel Memory Hierarchies," Workshop on Parallel I/O and Databases (DAGS 93), Dartmouth Institute for Advanced Graduate Studies, Hanover, NH.
- "Average-Case Analysis of Prediction" Dagstuhl-Seminar on Average-Case Analysis of Algorithms, Schloß Dagstuhl, Germany.
- "Paradigms for Optimal Sorting and Computational Geometry in Large-Scale Parallel Memories," Max Planck Institute, Saarbrücken, Germany.
- "Models for Parallel Secondary and Hierarchical Storage," Workshop on Models, Architectures, and Technologies for Parallel Computation, DIMACS, Rutgers University, New Brunswick, NJ.
- "Load Balancing Paradigms for Optimal Use of Parallel Disks and Parallel Memory Hierarchies," Stanford University, Stanford, CA.
- "Optimal Prediction via Data Compression," University of Texas at Dallas, Dallas, TX.
- "Load Balancing Paradigms for Optimal Use of Parallel Disks and Parallel Memory Hierarchies," Keynote address at Workshop on Algorithmic Research in the Midsouthwest (WARM 93), University of North Texas, Denton, TX.
- "Predictive Techniques for Caching and Locality Management," Microsoft Corporation, Redmond, WA.
- "Efficient Processing of Large-Scale Data," Mathematisches Forschungsinstitut Oberwolfach, Germany.
- "How to Predict Well," Tulane University, New Orleans, LA.
- "How to Predict Well," Supercomputing Research Center, Bowie, MD.

Dr. Paul G. Howard:

- nominee of the Department of Computer Science at Brown University for the ACM Doctoral Dissertation Award, based on work supported by this grant.

Prof. Philip M. Long:

- Invited talks:
    - "Simulating Access to Hidden Information while Learning," Australian national University Systems Engineering Seminar.

6

Duke University
DURHAM
NORTH CAROLINA
27708-0129

DEPARTMENT OF COMPUTER SCIENCE
BOX 90129

(919) 660-6500

December 19, 1994

Major David r. Luginbuhl
AFOSR/NM
110 Duncan Avenue, Suite B115
Bolling Air Force Base
Washington, DC 20332-0001

RE: AFOSR Grant No. F49620-92-J-0515

Dear Major Luginbuhl:

Per your conversation with Dr. Jeffrey Vitter, enclosed is the final project report for the above referenced grant. We apologize for the delay of this report and hope that it has not caused any inconvenience.

If you have any questions or concerns, please do not hesitate to contact us.

Sincerely,

Tina Gaither
Grants/Contracts Specialist

Enclosure

cc: E. Charniak
    L. Rossi